4Б

УДК 004.415.24 (004.932)

С.Л. Бедратюк
Київський національний університет імені Тараса Шевченка, м. Київ, Україна
stepan.bedratiuk@ukr.net

Стійкий до активних атак алгоритм постановки цифрових водяних знаків

У статті наводяться основні відомості про цифрові водяні знаки, приклади їхнього використання, проблеми, з якими стикаються розробники алгоритмів для вкраплення ЦВЗ у зображення. Також був запропонований алгоритм приховання ЦВЗ у зображення, який базується на підході компенсації геометричних спотворень і використанні точкових особливостей зображення, знайдених за допомогою SIFT-перетворення.

Вступ

Швидкий розвиток сучасних інформаційних технологій привів до того, що більшість продуктів інтелектуальної праці людей зберігається на цифрових носіях інформації. Вони не тільки зберігаються, але й розповсюджуються та продаються за допомогою Internet-комунікацій. Це привело до того, що виникла необхідність для захисту інтелектуальної власності, розміщеної у цифровому вигляді. Одним з найбільш ефективних способів вирішення цієї проблеми є використання цифрових водяних знаків (ЦВЗ) [1]. Використання терміна «водяні знаки» походить від поняття розміщення видимих водяних знаків на папері.

ЦВЗ для захисту авторських прав зазвичай використовують наступним чином:

1. Власник вкрапляє ЦВЗ в предмет інтелектуальної власності (зображення, відео чи аудіофайл) за допомогою спеціалізованого програмного забезпечення. ЦВЗ може містити інформацію, яка ідентифікує власника (ідентифікаційний код чи ПІБ), або ж детектор ЦВЗ буде фіксувати наявність чи відсутність ЦВЗ.
2. Власник розповсюджує файл за своїм бажанням.
3. При виникненні майнового конфлікту власник зможе довести свої права власності в суді, довівши наявність ЦВЗ у файлі.

Поряд з тим виділяють ще декілька галузей, в яких можуть бути використані ЦВЗ:

– Перевірка цілісності даних: додавання ЦВЗ до сигналу з камери відеонагляду (що дозволить потім довести (чи, навпаки, заперечити), що дані не були модифіковані сторонніми особами). Особливо актуальним такий захист може бути в сучасних засобах телефонного зв'язку – VoIP телефонії. Також цей спосіб може бути використаний для підтвердження цілісності знімків камер спостереження ДАІ.

– Прихована анотація документів: у медичні знімки можуть вкраплювати ЦВЗ з інформацією про пацієнта. Також додаткова інформація може додаватись до цифрових карт.

– Захист від копіювання: американські вчені запропонували метод, який зможе ідентифікувати людину, яка незаконно робила запис фільму в кінотеатрі. За цією схемою кожен кінотеатр отримував:

1. Фільм, з унікальним для цього кінотеатру ЦВЗ, вкрапленим у відеосигнал.
2. У кожну звукову доріжку вкраплювалися окремі ЦВЗ. Аналізуючи піратську версію фільму з відеосигналу вилучалися ідентифікатор кінотеатру. А конкретне місцезнаходження порушника в кінотеатрі визначалося з комбінації ЦВЗ на звуковій доріжці (точність складала декілька метрів).

– Прихована передача інформації: ЦВЗ можуть розглядатись як приховані повідомлення і використовуватись для таємних комунікацій (класична задача стеганографії) [1].

ЦВЗ (digital watermarking) як складова частина стеганографії розвивається дуже активно останнім часом. Ця дисципліна перебуває на перетині багатьох галузей сучасної математики – обробки сигналів, розпізнавання образів, теорії зв'язку, криптографії. Виходить велика кількість статей, присвячених цій проблематиці у світі.

Як бачимо, розвиток ЦВЗ повинні дуже актуальною темою в сучасній науці і має чіткий прикладний характер.

Основні характеристики цифрових водяних знаків

Загальну схему роботи з ЦВЗ для зображень можна зобразити наступною діаграмою (рис. 1):

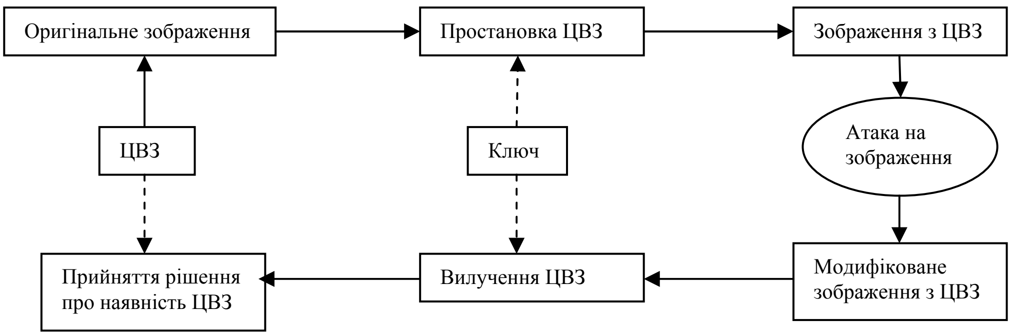


Рисунок 1 – Загальна схема роботи ЦВЗ для зображень

Пунктирні лінії не є обов'язковими складовими схеми (залежить від сценарію використання).

Як було описано у вступі, розрізняють декілька варіантів використання ЦВЗ. Для кожного з них ЦВЗ має мати різні характеристики. Розглянемо основні з них:

1. **Видимість (Perceptibility).**

a) Видимі ЦВЗ (perceptible): при видимому водяному знаку інформацію видно на фотографії або відео. Як правило, це інформація у вигляді тексту або логотипу, який ідентифікує власника. Наприклад, телеканал додає свій логотип на відео.

b) Невидимі ЦВЗ (imperceptible): при невидимому водяному знаку, інформація додається як цифрові дані до зображення, відео- або аудіофайла, але при цьому вона залишається непомітною для органів сприйняття людини.

2. **Стійкість (Robustness).**

a) **Крихкі ЦВЗ (fragile):** особливістю крихких ЦВЗ є те, що при найменших змінах контейнера ЦВЗ руйнуються. Саме вони використовуються для перевірки цілісності контейнера.

b) **Напівкрихкі ЦВЗ (semi-fragile):** цей тип ЦВЗ є стійким до незначних змін, (наприклад, до шумів), але руйнуються при більш серйозних спотвореннях контейнера.

c) **Стійкі ЦВЗ (robust):** ЦВЗ називається стійким, якщо його неможливо вилучити після застосування певного класу спотворень до контейнера.

3. **Пропускна здатність (Capacity):** характеризує довжину повідомлення, яке може бути прихованим у контейнері. Є два основні підходи:

a) одиничної довжини (1-bit watermark): у цьому випадку завдання детектора – визначати, чи містить контейнер ЦВЗ чи ні. Тобто передається лише один біт інформації;

b) при іншому підході визначається максимально можлива довжина ЦВЗ, яка може бути вкраплена в контейнер, не пошкодивши якість зображення.

4. **Складність алгоритму (Algorithm complexity):** різні схеми використання ЦВЗ накладають певні обмеження на складність обчислень, пов'язаних з вкрапленням і виявленням ЦВЗ. Наприклад, в сучасних DVD-програвачах є вбудована апаратна реалізація детектора ЦВЗ, тому час і необхідні ресурси для виявлення ЦВЗ мають бути мінімальними.

5. **Можливість видалення ЦВЗ (Reversible):** ця властивість показує, чи можливо відновити оригінальний контейнер після вкраплення ЦВЗ.

6. **Спосіб простановки ЦВЗ (Embedding method):**

a) модифікація в частотній області контейнера (spread-spectrum): характеризується більшою стійкістю, але має малу пропускну здатність. Найчастіше використовують перетворення Фур'є чи вейвлет-перетворення для переходу в частотну область;

b) модифікації в просторовій області контейнера (amplitude modulation): має більшу пропускну здатність, але є менш стійким. Найвідоміший метод – LSB (Least significant bit).

Активні атаки на ЦВЗ

Активними атаками називаються будь-які модифікації контейнера з ЦВЗ. Вони можуть бути як природними (наприклад, пошкодження при передачі через канали зв'язку), так і спеціально задіяні зловмисниками, щоб вилучити ЦВЗ.

При розробці алгоритмів мають враховуватись можливі модифікації контейнера. Оскільки в цій роботі розглядається тільки приховання ЦВЗ у зображення, то, відповідно, буде робитись акцент на пошкодженні зображень. Можна виділити наступні пошкодження:

1. Зображення можуть піддаватись компресії (наприклад JPEG, TIFF, GIF, PNG).
2. Геометричні перетворення:
   a) зсув зображення (translation);
   b) масштабування (scaling);
   c) поворот (rotation);
   d) обрізка (cropping);
   e) комбінація цих та інших геометричних перетворень.
3. Накладення шумів (як в частотну область зображення, так і в просторову).
4. Зміна контрасту.
5. Зміна кольорової гами.

Саме на ці атаки орієнтувався алгоритм, запропонований у цій роботі.

Алгоритм простановки ЦВЗ

Під дією геометричних атак відбувається десинхронізація ЦВЗ. На сьогоднішній день є два загальні підходи до вирішення проблем десинхронізації ЦВЗ. Перший підхід – компенсація геометричних перетворень перед вилученням ЦВЗ (тобто відбувається попередня обробка контейнера і його повертають до вигляду, в якому він був до атак). Другий підхід – це вкраплення ЦВЗ в область, інваріантну до геометричних перетворень [2].

Запропонований алгоритм використовує підхід компенсації геометричних спотворень на базі точкових особливостей контейнера. Точкові особливості знаходяться за допомогою алгоритму SIFT (Scale-invariant feature transform). Цей алгоритм є інваріантним до геометричних перетворень. Після виявлення ключових точок відбувається вкраплення бітів ЦВЗ в частотній області зображення (до якої ми переходимо після застосування дискретного косинус-перетворення).

Основні складові

Алгоритм використовує двохвимірний варіант **дискретного косинус-перетворення.** Його можна задати наступною формулою:

$$X_{k_1,k_2} = \sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} x_{n_1,n_2} \cos\left[\frac{\pi}{N_1}\left(n_1+\frac{1}{2}\right)k_1\right]\cos\left[\frac{\pi}{N_2}\left(n_2+\frac{1}{2}\right)k_2\right].$$

Іншою складовою частиною роботи є алгоритм з виявлення точкових особливостей **SIFT.** SIFT був вперше запропонований Ловом (David Lowe) в [3]. Метод SIFT дозволяє знаходити набір особливостей зображень, інваріантних відносно позиції, повороту та масштабування, та забезпечує стійкість до широкого спектра афінних спотворень, змін в рівні шуму, освітленості.

Кроки алгоритму постановки ЦВЗ

Крок 1. На вхід подається зображення *C* і інформація, яку потрібно приховати, *S* (у вигляді масиву бітів), а також ініціалізатор генератора псевдовипадкових чисел *A*.

Крок 2. Для того щоб знайти найбільш надійні точкові особливості зображення, робляться наступні дії:

1. До контейнера застосовують найбільш вірогідні геометричні спотворення і результат зберігають як окремі контейнери:

   – Збільшують масштаб, зменшують масштаб, роблять різні повороти зображення.

   – У результаті отримують масив контейнерів – *I[i]*.

   – Алгоритмом SIFT знаходяться точкові особливості для кожного контейнера *I[i]* окремо – *FeaturePoints[Image[i]]*.

2. Аналізуємо знайдені точкові особливості *FeaturePoints*[*I*[*i*]] і знаходимо тільки ті, які є спільними для всіх контейнерів – *CommonFP*.

3. Випадковим чином вибираємо 3 точкові особливості {FP1, FP2, FP3} з множини *CommonFP* і заносимо їх до ключа K.

Крок 3. На цьому кроці переходимо до частотного представлення зображення:

1. Розбиваємо зображення на блоки розміром 32×32 (можна робити розбиття і на 8×8 чи 16×16).
2. До кожного блоку застосовуємо ДКП.
3. У результаті отримаємо зображення в частотному представленні – *dctI.*

Для підвищення надійності блоки, які прилягають до границі зображення, не будуть використовуватись для вкраплення ЦВЗ (вони можуть бути сильно пошкоджені при поворотах та обрізці країв).

Крок 4. Генеруємо послідовність випадкових позицій, куди будуть вкраплюватись біти ЦВЗ – $\{x_i, y_i\}$, $i = 1..N$, де N – кількість бітів в ЦВЗ. Генератор ініціалізуємо за допомогою параметра *A*. Усі $\{x_i, y_i\}$ мають бути в інтервалі, який відповідає середнім частотам в блоках 16×16. На практиці бралися значення, які знаходились в діапазоні від 5 до 10. Ми не можемо модифікувати позиції, які знаходяться в лівому верхньому куті блоку, оскільки ми значно змінимо зображення. Також ми не можемо використовувати позиції, які знаходяться близько до інших кутів, оскільки вони можуть бути затерті/пошкодженні при стисканні інформації, добавленні шумів.

Крок 5. Процес вкраплення бітів буде складатись з наступних кроків:

Для кожного $\{x_i, y_i\}$, *S[i]* виконуємо наступні операції

1. Беремо черговий блок 16×16 з *dctI* – *block[i]*;

2. Виділяємо матрицю 3×3 (для більшої надійності можна брати більші розміри квадрата), яка оточує точку *$\{x_i, y_i\}$* в *block[i]*:

$$\begin{bmatrix} \{x_{i-1}, y_{i-1}\} & \{x_i, y_{i-1}\} & \{x_{i+1}, y_{i-1}\} \\ \{x_{i-1}, y_i\} & \{x_i, y_i\} & \{x_{i+1}, y_i\} \\ \{x_{i-1}, y_{i+1}\} & \{x_i, y_{i+1}\} & \{x_{i+1}, y_{i+1}\} \end{bmatrix}$$

3. Знаходимо мінімальне – *min*, максимальне – *max*, середнє значення по всіх елементах – *average* матриці;

4. Якщо S[i] = 1, тоді *max* – збільшуємо на величину
max = max + ||max-average| – |average-min|| + α, де α – параметр алгоритму.

5. Якщо S[i] = 0 тоді min – зменшуємо на величину;
min =min + ||max-average|-|average-min|| – α.

Крок 6. Застосовуємо обернене ДКП до кожного блоку dctI. Отримаємо W.

Крок 7. Формуємо ключ K. До ключа заносимо наступні параметри:
1. *{FP1, FP2, FP3} – точкові особливості;*
2. *N* – кількість бітів в ЦВЗ;
3. *A – ініціалізатор генератора псевдовипадкових чисел.*

Крок 8. W містить цифровий водяний знак і готовий до розповсюдження. ЦВЗ можна вилучити за допомогою ключа K.

Алгоритм вилучення ЦВЗ

Під час вилучення ЦВЗ знову знаходяться точкові особливості модифікованого контейнера і порівнюються з тими, які передані за допомогою ключа. Після цього відбувається компенсація геометричних спотворень і вилучення бітів ЦВЗ.

Крок 1. На вхід подається *ChangedW* (ми припускаємо, що зображення піддавалось модифікації) *і* ключ *K.*

Крок 2. Застосуємо алгоритм SIFT для виявлення точкових особливостей зображення – *FeaturePoints[ChangedW].*

Крок 3. Витягаємо з *K* характеристики точкових особливостей – *{FP1, FP2, FP3}.*

Крок 4. Знаходимо в *FeaturePoints[ChangedW]* точкові особливості, які відповідають *{FP1, FP2, FP3}* базового зображення – *{wFP1, wFP2, wFP3}.*

Крок 5. Три точкові особливості з ключа фактично задають трикутник – *{FP1, FP2, FP3}*, а точкові особливості зображення задають модифікований трикутник (після геометричних спотворень) – *{wFP1, wFP2, wFP3}*. Тобто, щоб відновити зображення, нам потрібно застосувати ті самі операції повороту/зміни масштабу, які потрібні для переведення трикутника *{wFP1, wFP2, wFP3}* в трикутник *{FP1, FP2, FP3}*. Задіємо ці ж операції до *ChangedW* і отримаємо зображення *W.*

Крок 6. Застосовуємо ДКП до блоків *W.* Аналогічно до того, як ми це робили при прихованні ЦВЗ. В результаті отримаємо *dctW.*

Крок 7. Витягнемо з ключа *A* та *N* – і генеруємо *N* позицій, в які були приховані біти ЦВЗ – *$\{x_i, y_i\}$* за тим же принципом, що й при прихованні.

Крок 8. Процес виявлення бітів буде складатись з наступних кроків:

Для кожного *$\{x_i, y_i\}$*, виконуємо наступні операції:

1. Беремо черговий блок 16×16 з *dctI – block[i];*
2. Виділяємо матрицю 3×3, яка оточує точку *$\{x_i, y_i\}$* в *block[i]*:

$$\begin{bmatrix} \{x_{i-1}, y_{i-1}\} & \{x_i, y_{i-1}\} & \{x_{i+1}, y_{i-1}\} \\ \{x_{i-1}, y_i\} & \{x_i, y_i\} & \{x_{i+1}, y_i\} \\ \{x_{i-1}, y_{i+1}\} & \{x_i, y_{i+1}\} & \{x_{i+1}, y_{i+1}\} \end{bmatrix}$$

3. Знаходимо мінімальне – *min*, максимальне – *max*, середнє значення по всіх елементах – *average;*
4. Якщо |*max-average*|>|*average-min*|, *тоді S[i]=1 інакше S[i]=0.*

Крок 9. Якщо витягнуте ЦВЗ збігається з прихованим, то ми можемо стверджувати, що до нас потрапило саме те зображення, яке ми розповсюдили. На практиці можна приймати зображення за оригінальне за умов, що витягнуте ЦВЗ наближається до прихованого і рівень різниці між ними не перевищує β.

Аналіз якості алгоритму

Продемонструємо роботу алгоритму. Як контейнер *I* візьмемо наступне зображення (розміром 512×512) (рис. 2).



Рисунок 2

Як ЦВЗ візьмемо масив випадкових бітів довжиною N = 100.

Застосуємо поблокове ДКП до початкового зображення і виконаємо вкраплення бітів згідно з вищеописаним алгоритмом.

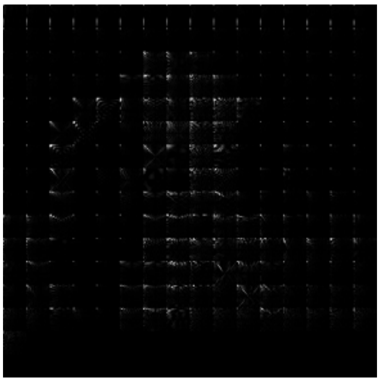


Рисунок 3 – Поблокове застосування ДКП



Рисунок 4 – Зображення з вкрапленим ЦВЗ

Зробимо аналіз спотворень, які були завдані початковому зображенню. Для цього були взяті наступні критерії [4] (табл. 1).

Як бачимо, тільки перший критерій дав відносно велике відхилення. Це був прогнозований результат, оскільки при прихованні кожного біта ми фактично збільшували максимальне значення блоку 3×3 як мінімум на величину α.

Критерії, які показують відносну зміну зображення, дають досить малу різницю між двома зображеннями.

Знайдемо ключові точки зображення (вони представлені у вигляді векторів на рис. 5 і рис. 6) і для прикладу продемонструємо спільні ключові точки між початковим зображенням і зменшеним в масштабі та повернутим на 90 градусів.

Таблиця 1

Критерій	Значення на I	Значення на W
Максимальна різниця: $\max_{x,y} \left| I_{x,y} - W_{x,y} \right|$	0	11,0000
Якість зображення: $1 - \frac{\sum_{x,y}(I_{x,y} - W_{x,y})^2}{\sum_{x,y}(I_{x,y})^2}$	1	0,999991
Структурний вміст: $\frac{\sum_{x,y}(I_{x,y})^2}{\sum_{x,y}(W_{x,y})^2}$	1	0,9998



Рисунок 5 – Ключові точки зображення

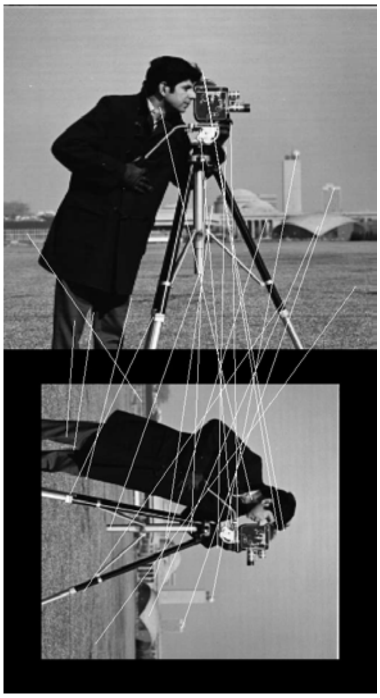


Рисунок 6 – Спільні ключові точки

Тепер модифікуємо зображення, яке містить ЦВЗ: змінимо яскравість, зменшимо масштаб, повернемо зображення на 180 градусів. За допомогою ключа відновимо початкове зображення (рис. 7 та рис. 8).



Рисунок 7 – Модифіковане зображення з ЦВЗ



Рисунок 8 – Зображення після реконструкції

Слід відзначити, що зміна яскравості і збереження зображення в формі JPEG (тобто стиснення) практично не впливають на процедуру вилучення ЦВЗ завдяки схемі вкраплення.

Напрямки подальших досліджень

До недоліків можна віднести наступні речі:

1. Не у всіх випадках вдавалось коректно відновити зображення після геометричних перетворень. Щоб покращити ситуацію, потрібно брати не 3 ключові точки, а більшу кількість. Тоді відновлення може бути більш достовірним.

2. Якщо до зображення застосовувалось геометричне перетворення, то не вдавалось вилучити ЦВЗ з високим рівнем збігу з початковим (>90%). З іншого боку, якщо не було задіяно геометричних перетворень, а тільки стиснення і зміна яскравості, то вдавалось відновити ЦВЗ з рівнем надійності >95%.

3. Алгоритм SIFT дає хибну ключову точку з ймовірністю 5%. Тобто може виникнути ситуація, коли ми виберемо не ключову точку. Саме для цього був доданий Крок 2 алгоритму вкраплення. Але це все одно не дає повної гарантії.

4. Відносно великий час роботи алгоритму вкраплення (хоча загальна складність складає *O(N*Log(N)*), де N – кількість пікселів у зображенні). Найбільше часу йшло на виявлення стійких точкових особливостей на Кроці 2.

5. Невелика пропускна здатність алгоритму – в найкращому випадку можна приховати не більше 10% інформації від початкового розміру зображення.

При реалізації програми використовувалась бібліотека OpenCV (для обробки зображень) та бібліотека Роб Хесса(Rob Hess) для роботи з SIFT.

Висновки

В роботі подані основні відомості про цифрові водяні знаки, схеми їх використання, основні характеристики. Також наведений алгоритм вкраплення невидимого ЦВЗ, стійкого до геометричних спотворень, стиснення та зміни яскравості. Наведено ряд переваг та недоліків цього методу та шляхи усунення недоліків. Слід відзначити, що на даний момент ще не розроблено алгоритм, який би повністю задовольняв всім критеріям стійкості.

Література

1. Грибунин В.Г. Цифровая стеганография / Грибунин В.Г., Оков И.Н., Туринцев И.В. – Солон-Пресс, 2002. – 265 с.
2. Кошкина Н.В. Методи синхронізації цифрових водяних знаків / Н.В. Кошкина // Кібернетика і системний аналіз. – 2008. – № 1. – С.180-188.
3. Lowe D.G. Distinctive image features from scale-invariant keypoints / D.G. Lowe // International Journal of Computer Vision. – 2004. – № 2. – P. 91-110.
4. Швидченко И.В. Стойкие криптостеганографические алгоритмы / И.В. Швидченко // Искусственный интеллект. – 2009. – № 1. – С. 218-226.

S.L. Bedratiuk

Watermark Algorithm Robust to Active Attacks

Article contains the basics of digital watermarking, examples of their use, the problems faced by developers of watermark algorithms. Also it is proposed robust watermark algorithm for the picture containers, which is based on geometric distortion compensation approach and uses image’s point features founded using SIFT.

Стаття надійшла до редакції 06.07.2010.